Volume 13 Number 24 1985 Nucleic Acids Research

Putative amino acid sequence of chick calcium-binding protein deduced from a complementary DNA sequence

P.W.Wilson, M.Harding and D.E.M.Lawson

AFRC Institute of Animal Physiology, Babraham, Cambridge CB2 4AT, UK

Received 24 October 1985; Revised and Accepted 22 November 1985

ABSTRACT

Two DNA fragments coding for chick CaBP have been isolated and sequenced. cDNA was prepared from enriched intestinal mRNA and cloned in pUC12[32]. The recombinant clones were screened by differential hybridisation with ^{32}P-cDNA probes synthesized from vitamin D replete and deficient chick intestinal mRNA. Two clones had outstanding affinity with the +D probe. Hybrid-arrested and hybrid-selected translation systems showed that both clones hybridised to mRNA coding for immunoprecipitable CaBP. The mRNA for CaBP has a 100 bp G,C rich sequence before a 786 bp coding region followed by 1250 nucleotides 3' untranslated region. Nucleotides coding for the Ca-binding sites show a high degree of homology for Ca-binding sites in chick calmodulin and rat intestinal CaBP. The amino acid sequence specified by the longest open reading frame contains five Ca-binding sites but is too large for the native CaBP; post-translational modification must therefore occur.

INTRODUCTION

CaBP occurs in two forms; the larger one, M_r=27500 is found in all tissues except muscle and mammalian intestine, whereas the smaller, M_r=8600, is found only in mammalian intestine and kidney (1). The highest concentration of these two proteins is found in the intestine and a variety of observations have implicated them in vitamin D stimulated Ca absorption. The amino acid sequence of the smaller protein has been established (1,2) but since antisera raised against one form does not cross react with the other it is probable that this sequence is quite distinct from that of the larger CaBP. Chick intestine contains the larger CaBP, M_r=27500, and has proved a convenient system to investigate the molecular action of vitamin D and the biochemical steps in Ca translocation across mucosal cells. Thus the synthesis of CaBP mRNA in chick intestine is dependent upon the hormonal form of vitamin D, 1,25-$(OH)_2$D, and in the hormone's absence the synthesis quickly stops (3). This effect of 1,25-$(OH)_2$D occurs in the nuclei of the mucosal cells. Although partial cDNAs for CaBP have been

© IRL Press Limited, Oxford, England.

prepared (4-6) a full length cDNA for the larger protein would be useful for investigating a number of problems. In this case the amino acid sequence of the protein, at present unknown, could be deduced and this information might stimulate further experiments into the molecular events in which it is involved. The cDNA could explain the unexpected length of CaBP mRNA (>2000 nucleotides) compared to the size of the protein, about 240 amino acids.

MATERIALS AND METHODS

Materials Nuclease-treated rabbit reticulocyte lysate was purchased from Dr. R.J. Jackson (University of Cambridge, Department of Biochemistry). Vanadyl ribonucleoside complex, oligo (dT) cellulose, DNA polymerase I, RNase H, T_4 DNA polymerase and bacterial alkaline phosphatase were from Bethesda Research Laboratories (UK) Ltd. [^{32}P]dCTP (3000 Ci/mmol), [^{35}S]dATP (410 curies/mmol) and L-[^{35}S] Methionine (1040 curies/mmol) were obtained from Amersham International. Reverse transcriptase was purchased from Anglian Biotechnology Ltd. T_4DNA ligase and calf intestinal alkaline phosphatase were from Boehringer Mannheim. Oligo $(dT)_{12-18}$ primer was from P.L. Biochemicals Inc. RNasin was from P & S Biochemicals Ltd. M13 single strand primer and DNA polymerase 1 (Klenow Fragment) were from Pharmacia. Nitrocellulose filters BA-85 (0.45μm) were from Schleicher & Schuell. Restriction endonucleases were purchased from Pharmacia, P & S Biochemicals Ltd., Bethesda Research Laboratories (UK) Ltd. and Boehringer Mannheim.

Animals One day old White Leghorn chicks were fed on a cholecalciferol deficient diet (7) and used after four to five weeks. Vitamin D dosed chicks received 500 i.u. cholecalciferol orally in 0.1ml 10% ethanol, 90% propylene glycol 72 h before killing.

Preparation of CaBP mRNA Polysomes were prepared from chick ileal mucosa as described previously (8) except that the ribonuclease inhibitor vanadyl ribonucleoside complex was included at a level of 10 mM in the homogenisation buffer in the place of rat liver cell sap. Poly (A^+) RNA was then prepared by chromatography on oligo (dT) cellulose (9). Poly(A^+)RNA from vitamin D replete birds is referred to as +D mRNA and that from vitamin D-deficient birds as -D mRNA.

Sucrose gradient enrichment of CaBP mRNA A 3-4 fold enriched CaBP mRNA (18S) was obtained by centrifugation of formamide treated +D mRNA on linear 5-20% sucrose gradients prepared in 1 mM EDTA, 10 mM Hepes pH 7.5. Centrifugation was for 20h at 35,000 rpm in a Beckman SW 40 rotor at 4°C

(10). Gradient fractions were analysed for CaBP mRNA by translation using L-[^{35}S] methionine in the nuclease-treated rabbit reticulocyte lysate (11) followed by immunoprecipitation of radioactive products with CaBP antiserum.

Synthesis of blunt-ended ds cDNA The method of Gubler and Hofmann (12) was followed using the RNase H-DNA polymerase 1-mediated second strand synthesis.

The first strand cDNA was synthesised in a volume of 50µl containing 50 mM Tris-HCl pH 8.3, 10 mM $MgCl_2$, 10 mM DTT, 4 mM Na-Pyrophosphate, 1.25 mM dNTP's, 100 µg/ml oligo $(dT)_{12-18}$, 40 µg/ml actinomycin D, 600 units/ml RNasin, 60 µg/ml +D mRNA (preheated to 70°C 1-2 min., then quick chilled), 2,200 units/ml reverse transcriptase, for 1 h at 43°C. The addition of RNasin ribonuclease inhibitor was found to greatly increase the proportion of full length transcripts obtained. The products were phenol extracted and passed down a spin column containing 0.9 ml Sephadex G-50 equilibrated with 5 mM Tris-HCl pH 7.5.

Second strand synthesis was carried out in 100µl of 20 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$, 100 mM KCl, 50 µg/ml bovine serum albumin, 40 µM dATP, 40 µM dGTP, 40 µM dTTP, 30 µM dCTP, 2.5 µCi [α-^{32}P]-dCTP, 12 units/ml of E. coli RNase H, 320 units/ml DNA polymerase 1. Sequential incubations of 1 h at 12°C and 1 h at 22°C were carried out. The products were extracted twice with phenol and then passed down a spin column equilibrated with 10 mM Tris-HCl pH 7.5.

The size of the transcripts produced was checked by agarose gel electrophoresis and corresponded to an average length of 2 kb. This cDNA was then treated with T_4 DNA polymerase to produce ends suitable later for blunt end ligation into pUC. Incubations in a volume of 50µl containing 10 mM Tris-HCl pH 8.5, 5 mM $MgCl_2$, 14 µM dNTP's, 200 units/ml T_4 DNA polymerase, 3 µg/ml ds cDNA were carried out at 15°C overnight. The products were phenol extracted and passed through a spin column equilibrated in 10 mM Tris-HCl pH 7.5, 1 mM EDTA. A final yield of 233 ng blunt ended ds cDNA was obtained from 2.8 µg + D mRNA (gradient enriched).

Preparation of [^{32}P] cDNA probes Gradient purified 18S +D and -D mRNA's were used to synthesise ds cDNA as described above. Labelling of the cDNA was carried out using the Amersham Nick Translation Kit and [^{32}P] dCTP label. Specific activities of 3-4 x 10^8 dpm/µg DNA were obtained.

Construction of recombinant plasmids by blunt end ligation The cloning vector pUC12 was digested with Sma I, then purified on a sucrose gradient

(13) to obtain a low transformation background and subsequently dephosphorylated by bacterial alkaline phosphatase at 65°C for 2 h. 20 ng of blunt ended ds cDNA were incubated with 40 ng of vector in a volume of 10 µl containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 2mM ATP, 500 µg/ml bovine serum albumin, 90 units/ml T_4 DNA ligase at 15°C overnight.

Transformation of E. coli by recombinant plasmids Recombinant plasmids were used to transform E. coli DHI by the high efficiency method described by Hanahan (14).

High-density screening of transformants The colonies were grown at 30°C on nitrocellulose filters (BA-85) placed on agar plates containing ampicillin (50µg/ml). Filters containing colonies were blotted to produce duplicates for hybridisation with each of the +D and -D cDNA probes. The master filter was grown on an agar plate containing 5% glycerol for 1 h at 37°C and then after covering with a second nitrocellulose filter was sandwiched between Whatman filters, sealed in a plastic bag and stored at -70°C (15). Following the lysis of colonies, in-situ differential hybridisation was carried out overnight at 68°C, using 1-2 ng (2.5 x 10^5dpm) of +D or -D cDNA probes /ml of hybridisation buffer, (16). After stringent washing radioactive colonies were detected on the dried filters by autoradiography.

Dot blot analysis of recombinant DNA Denatured plasmid DNA from selected clones was bound to nitrocellulose filters (17). Hybridisation using +D and -D [^{32}P] cDNA probes was carried out as described above.

Hybrid-arrested translation Ten micrograms of plasmid DNA prepared by the cleared lysate, CsCl method was linearised by Xba I digestion and then added to 5 µg +D intestinal mRNA in 10mM Pipes, pH 6.4, 0.4 M NaCl, 80% deionised formamide in a final volume of 25µl. Following heat denaturation at 75°C for 15 min, the samples were incubated for 2h at 48°C. After hybridisation, the samples were diluted with yeast t-RNA, (25µl/200µl); 100µl was left in hybrid form, whilst the other 100µl was heated for 1 min at 100°C and then quick-chilled to release the hybridised mRNA (18). After ethanolprecipitation the samples were disolved in water (5µl) and then translated in the nuclease-treated rabbit reticulocyte lysate system. Samples were taken for immunoprecipitation with CaBP antiserum and trichloroacetic acid precipitation (10). The radioactivity in the immunoprecipitates was measuured directly following resuspension in 5% TCA and filtration on GF/C filters.

Hybrid selected translation Following linearisation by XbaI, 6µg of

plasmid DNA in water (7µl) was heated to 100°C for 10 minutes and chilled on ice. After denaturation in 0.5 M NaOH and neutralisation by the addition of 0.5 volumes of a solution of 1M NaCl, 0.3M sodium acetate, 0.5M Tris-HCl (pH 8.0) and 1M HCl, the DNA was spotted onto 4mm^2 nitrocellulose filters (Millipore HAWP) under slight vacuum (17). The filters were air dried, washed with 6 x SSC and baked for 2h at 80°C in a vacuum oven. Hybridisation was carried out in 65% deionised formamide, 20 mM Pipes pH 6.4, 0.2% SDS, 0.4 M NaCl, 100µg/ml calf liver tRNA and 100µg/ml +D mRNA in a final volume of 65µl. The solution was heated at 70°C for 10 min and then incubated at 50°C for 3.5 h. The filters were washed 10x with 1ml of 10mM Tris-HCl pH7.6, 0.15M NaCl, 1mM EDTA, 0.5% SDS at 65°C, and twice with the same buffer without SDS. To elute the hybridised mRNA, calf liver tRNA (15µg/150µl) was added to each filter and the tubes were placed in a boiling water bath for one minute and then quick frozen in liquid N_2 (19). Following ethanol precipitation the mRNA was dissolved in 5 µl water for translation in the nuclease-treated rabbit reticulocyte lysate. Immunoprecipitation, gel electrophoresis and processing were performed as before.

DNA sequencing Two protocols were used for this analysis. In the first, restriction enzyme digestion of clone pWH11 was performed followed by insertion into the appropriate sites of the vectors M13 mp 18 and mp 19 (which had been dephosphorylated using calf intestinal alkaline phosphatase) (20). Alternatively the shotgun cloning method described by Bankier & Barrell (21) was used. Sonication of WH11 to generate a large number of random fragments, end repair using T_4 DNA polymerase and size fractionation on an agarose gel were carried out. The size range of 300-700bp was selected and inserted into the Sma I site of mp18.

Transformation of competent JM101 cells by both sets of vectors was performed and white plaques (containing inserts) were removed and used for preparation of single stranded template DNA for sequence analysis. Sequencing reactions were carried out according to the Sanger dideoxy chain termination procedure using [^{35}S] dATP and the products analysed on 50 cm x 20 cm x 0.3 mm plates using buffer gradient gels (20-22).

RESULTS

Selection of chicken cDNA clones by differential hybridisation A partial cDNA library of about 2000 recombinant clones was constructed using 20.0 ng of cDNA prepared from enriched intestinal mRNA from vitamin D replete birds as starting material. After high density screening colonies were only

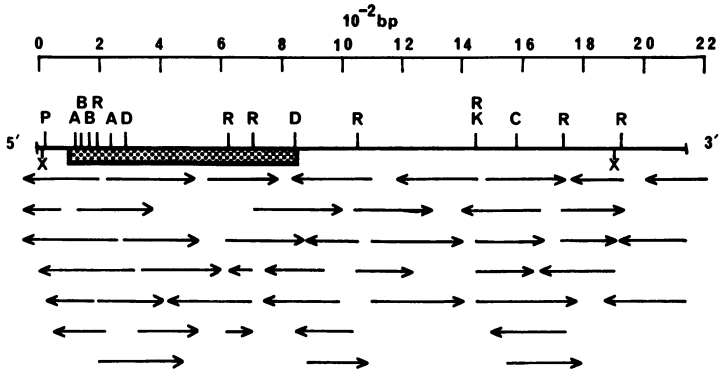

Fig.1 Restriction map of cDNA insert in pWH11 and sequencing strategy. Only the restriction sites relevant to the analysis of pWH11 and pWH12 and their sequencing are shown. The extent and direction of sequencing are indicated by arrows. pWH12 is orientated in pUC12 in the opposite direction to pWH11 and extends between the points marked x. Coding region is indicated by shaded area. A, Pst I, B, Bgl II; C, Hinc II; D, HinD III; P, Pvu I; R, Rsa I.

selected for further screening if a positive signal was obtained with the +D probe from colonies on both duplicate filters but no signal was obtained with the -D probe on either duplicate filter. Eight positive colonies were obtained as a result of this stringent procedure. These clones were mapped by restriction endonucleases to determine the size and orientation of the insert. For 6 of the plasmids the inserts ranged in size from 1.1 to 2.1 kbp with the remaining two inserts being 100 and 200 bp respectively. The plasmid DNA from the larger clones was re-screened on nitrocellulose filters using the dot-blot technique. Controls included in this screening were the pUC12 vector and three recombinant clones which had hybridised equally well to both +D and -D probes in the initial screening. In this system three of the clones pWH11, pWH12 and pWH17 showed a much greater hybridisation affinity for the +D cDNA probe compared to the -D cDNA probe, whereas pWH13 produced only a slightly increased hybridisation signal with the +D cDNA probe.

The colonies with the two largest inserts (designated pWH11 and pWH12) were chosen for further analysis. A partial restriction endonuclease map is given in Fig. 1; EcoR I, BamH I, Pvu I, Sac I, Sal I, Sma I, Xho I and Xba I do not cleave the insert in either clone. The pattern of enzyme digests given by pWH11 and pWH12 indicated that the inserts were present in

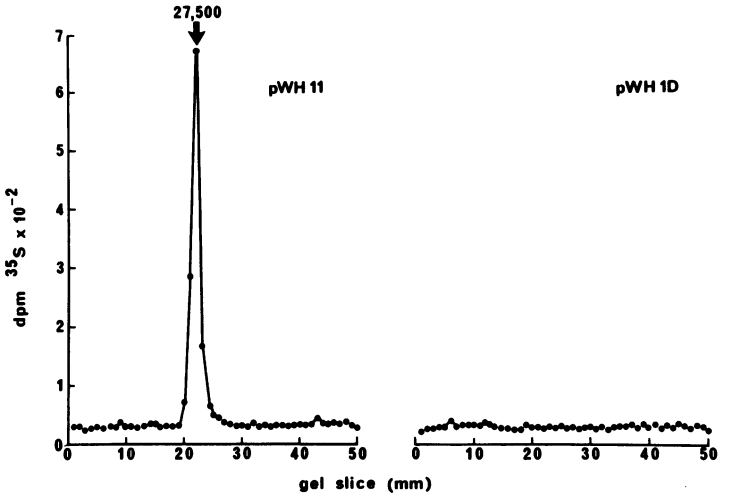

Fig.2 Identification of protein encoded by pWH11 and pWH12 by hybrid--selected translation system. The translation products of the hybridised mRNA were immunoprecipitated with anti-CaBP serum and, after solubilising, electrophoresed as described in the methods. pWH11 and pWH12 (data not shown) gave a single peak of labelled protein comigrating with authentic CaBP. Control clone pWH1D (i in Fig.1) gave no radioactive peak in the immunoprecipitate.

the opposite orientation and that pWH12 insert was about 200 bp shorter than pWH11.

Analysis of the mRNA hybridised by pWH11 using the hybrid-arrested translation system showed that this plasmid coded for CaBP. After hybridising total intestinal mRNA to pWH11 there was a 50% reduction of translation of CaBP mRNA before heat treatment of the hybrids, in contrast to no reduction with the control clone pWH1D, which hybridised with equal affinity to the ^{32}P-probes in the original screening.

The use of the hybrid-selected translation system led to the same conclusion. Fig. 2 shows that clones pWH11, pWH12 both hybridise to a mRNA which codes for immunoprecipitable CaBP, whereas the control recombinant plasmid pWH1D did not.

Sequence analysis of pWH11 The primary nucleotide sequence of pWH11 is shown in Fig. 3 together with the longest open reading frame beginning at the first ATG triplet from the 5' end. The length of the cDNA was 2136 bp but as this did not contain a poly (A) tract representing the original poly

```
   1 GGAGTCGTCGTCGCCGGAGCTGCCGCCGTCGCCGCTCCGGCCGCGGGCACAGTCAGCACCGCGGACAGCGCCC
                                  1                  10
                                  MetThrAlaGluThrHisLeuGlnGlyValGluIleSerAlaAla
  74 CCGCTGAGCCCCCTCTGCACCCCCATCATGACGGCGGAGACGCACCTGCAGGGCGTGGAGATCTCGGCCGCC
                      20                  30
     GlnPhePheGluIleTryHisHisTyrAspSerAspGlyAsnGlyTyrMetAspGlyLysGluLeuGlnAsn
 146 CAGTTCTTCGAGATCTGGCACCACTACGACTCCGACGGCAATGGGTACATGGATGGGAAGGAGCTACAAAAC
          40                  50                  60
     PheIleGlnGluLeuGlnGlnAlaArgLysLysAlaGlyLeuAspLeuThrProGluMetLysAlaPheVal
 218 ITCATCCAGGAGCTGCAGCAGGCGCGGAAGAAGGCAGGCTTGGACTTAACACCTGAAATGAAAGCTTTTGTG
                  70                  80
     AspGlnTyrGlyLysAlaTyrAspGlyLysIleGlyIleValGluLeuAlaGlnValLeuProThrGluGlu
 290 GACCAGTATGGCAAGGCCACTGATGGAAAAATAGGAATCGTTGAGCTTGCTCAGGTGTTGCCGACGGAGGAG
              90                  100                 110
     AsnPheLeuLeuPhePheArgCysGlnGlnLeuLysSerSerGluAspPheMetGlnThrTryArgLysTyr
 362 AATTTCCTGTTGTTCTTTAGGTGCCAGCAGCTAAAGTCAAGTGAAGACTTCATGCAGACATGGAGAAAATAT
                      120                 130
     AspSerAspHisSerGlyPheIleAspSerGluGluLeuLysSerPheLeuLysGluLeuLeuGlnLysAla
 434 GACAGTGACCACAGTGGTTTCATTGATTCTGAGGAACTTAAGAGTTTCTTGAAAGATTTATTACAGAAAGCA
          140                 150
     AsnLysGlnIleGluAspSerLysLeuThrGluTyrThrGluIleMetLeuArgMetPheGluAlaAsnAsn
 506 AATAAGCAGATTGAAGACTCAAAGCTAACAGAATATACAGAAATAATGCTCAGGATGTTTGATGCAAACAAT
          160                 170                 180
     GluGlyLysLeuGluLeuThrGluLeuAlaArgLeuLeuProValGlnGluAsnPheLeuIleLysPheGln
 578 GATGGAAAATTGGAGCTTACTGAACTGGCCAGGCTACTCCCAGTACAGGAAAATTTTCTTATTAAATTTCAG
                  190                 200
     GlyValLysMetCysAlaLysGluPheAsnLysAlaPheGluMetTyrAspGlnAspGlyAsnGlyTyrIle
 650 GGTGTCAAAATGTGTGCAAAAGAGTTCAATAAAGCCTTTGAGATGTACGATCAAGATGGCAATGGATATATA
      210                 220                 230
     AspGluAsnGluLeuAspAlaLeuLeuLysAspLeuCysGluLysAsnLysLysGluLeuAspIleAsnAsn
 722 GATGAAAATGAACTTGATGCACTACTGAAGGATCTCTGTGAAAAGAACAAAAAGGAATTAGACATTAACAAC
                      240                 250
     LeuAlaThrTyrLysLysSerIleMetAlaLeuSerAspGlyGlyLysLeuTyrArgAlaGluLeuAlaLeu
 794 CTTGCGACATACAAGAAAAGCATCATGGCCTTGTCTGATGGAGGGAAGCTTTACCGAGCAGAACTGGCTCTC
              260
     IleLeuCysAlaGluGluAsn
 866 ATTCTCTGTGCTGAGGAAAATTAAAACTCTTCTCTCATGTCCACTTAACTAGTGATGTATTCTATCTACACA

 938 ATAACTGTGCACTATAAGGGAGTAGGCTGTATTTTTAAACTGCATATAGAAAATTAGCCAGGATGTGTGGCA

1010 CATTCCTTTAAGTTTGTTTCTATACTGTTTGTAATGTACAGTTTTTGTAACAATAAGATTGATAAAGAGAAT

1082 GTCTATGTTTGGGCCAGTCTGTATATTCAAAAGAAACTAAAACATGTTGGGGTTGGATTTTTTTTTTTTTTT

1154 TGGCTTTTTTGGCTCTTTTTTTTTTGCTTTCATGAACATGCCTGAAAAAAAAATTACACCTGCTGATGTTGTG

1226 GTCATCATATCCTTTGACACTTGCAACATTTTTCCTTGTTGATCTACATAACAAAAGGCCTAGAAATCTCTC

1298 TCCCACTTACAACTACACACATTCGGGTTTTTTTGTTTGGGTTGTTGGTTTTTTTTTTTTAATGTATGATACA

1370 AATTCAGAAGAAATATGATTTAGGTTGTAATAGGATAAAGAAGTAGGCATGTTTACACTCCAAAAGGTACCG

1442 AGAAATATTTCCAGTTGTCAAACCACCCACTTACAGGTCTGTGCTCTCTGTAGTTTTTCAACTACTCTTCTC

1514 ATTATGAACTCAAAGTATATAGAGGAAGCATTTATCTGGAATAGAACTCCATGCAGTAGTTGACAAAAATAT

1586 ACTGTTCTCAAACTTGCTGTTTAACTTTACTGAATTTAAACATAGGCACTTCAGAAACAATGCCTTTAATCT
```

1658 GTCTTGGAATCCTGGCTAAGTGACAGATAGCATAGTTAATACACAGATTAATATATGATATTATAAGTGTAC

1730 CTTTCATGACTATTGCTGTGTCAGAGAATATGACAATCCATTTTCTAAACTATTTTCACATTTTGCAGGTTA

1802 TATTATTCTAGTAAATTGCTGTTTTTACATCATATTCTGTGTAACCTATAATTAAATTTAATATCCTAAGAC

1874 TATTGTTGTCTAGTTTGTCTATCTCCTGGATGCTCTTTTCTGTACCATGTAAAGGACAAAACAGTTATTTTG

1946 AAAAATTGTGCCTCCTTGGATCTTACACTGAGTTATTAATCTGTAATAATACTAATAAAAGGTAACACTAAA

2018 ATAAACACAAGACCTAGTAAAGATTTTATAACAGTTTAGTAATTCGGTTTACCAGTGCTCTGCTTTTTATTA

2090 AGCTAGAGAGGTATTTTTTAAAATGTGAAATATAATATGGAACAAGC

Fig.3 Nucleotide sequence of pWH11 and translation of longest open reading frame. The numbers at the side indicate nucleotide positions and those above the amino acid positions. The calcium binding regions are boxed. pWH12 has a 5'-end at position 9 and a 3'-end at position 1884 of pWH11.

(A^+) tail, the length of the mRNA for CaBP must be a little greater than this. The longest open reading frame of pWH11 begins from the first nucleotide but the first ATG initiation signal is found 101 nucleotides downstream. In this frame the first termination codon is found after 887 nucleotides giving a possible coding region of 786 nucleotides and a 3'-nontranslated region of 1250 nucleotides. The amino acid sequence of CaBP, M_r27500, can be deduced from the cDNA nucleotide sequence in Fig. 3. Comparison of the amino acid composition calculated from the open reading frame beginning at position 104 with that reported previously (23) from analysis of pure chick CaBP is given in Table 1. The deduced sequence has a substantially larger molecular weight than that estimated by gel filtration, SDS polyacrylamide gel electrophoresis and the amino acid composition (24). A second initiation codon is found at position 194 and the molecular size of this translation product (26,500) is in closer agreement with the measured values (Table 1). However, the values for two of the amino acids histidine and tryptophan differ by more than can probably be ascribed to experimental error. It would seem therefore that the N-terminal amino acid lies between these two extremes thereby allowing another tryptophan residue and two more histidines to be included in the amino acid composition without substantially affecting the proportion of the other amino acids.

Calcium Binding Domains Each molecule of chick CaBP has four high affinity Ca binding sites (23). Analysis of the crystalline structure of parvalbumin (25) led to the identification of the structure of these sites

Table 1
Comparison of the observed CaBP amino acid composition and those calculated from cDNA sequence

	Observed values[23]	Calculated values	
		CaBP-104	CaBP-197
MW	27500	30000	26500
cys	2	4	4
try	2	2	1
pro	3	3	3
his	4	4	1
val	5	6	5
arg	6	6	6
met	7	8	7
tyr	8	9	7
thr	9	10	8
ser	10	11	9
ile	11	13	11
gly	13	15	12
phe	13	15	13
ala	17	19	16
lys	23	25	25
leu	29	33	32
asp	31	34	31
glu	39	43	38

CaBP-104 and CaBP-197 refers to the open frame readings between nucleotides 104 to 886 and 197 to 886 respectively.

in proteins which bind Ca with a high affinity. The general form of the amino acid sequence of these domains (Table 2) has been deduced from analysis of several Ca binding proteins including parvalbumin, troponin C, calmodulin, brain S-100 protein. It consists of a sequence of 12 amino acids, five of which contain an oxygen function in the side chain, linking two helical regions each of 8 amino acids, of which half are hydrophobic. For those proteins in which Ca binding is critical for function, the domains responsible can be expected to be highly conserved. Comparison of this consensus sequence with the open reading frame sequence in Fig. 3 has enabled five putative calcium binding regions to be identified (Table 2) in the potential translation product of CaBP mRNA.

Table 3 shows the nucleotide homologies between those portions of chicken cDNA clones that correspond to the amino acids in the Ca-binding

Table 2

	E	Z	-	-	Z	Z	-	-	Z	O	-	O	-	O	G	-	I	O	-	-	O	Z	-	-	Z	Z	-	-	Z
calmodulin I	G	F	K	E	A	F	S	L	F	D	K	D	G	D	G	T	I	T	T	K	E	L	G	T	V	M	R	S	L
calmodulin II	E	L	Q	D	M	I	D	E	V	D	A	D	G	N	G	T	I	D	F	P	E	F	L	T	M	M	A	R	K
troponin II	E	L	D	A	I	I	E	E	V	D	E	D	G	S	G	T	I	D	F	E	E	F	L	V	M	M	V	R	Q
porcine CaBP I	E	L	K	G	I	F	E	K	Y	^{A}A	K	E	G	D	P	^{N}Q	L	S	K	E	E	L	K	L	L	L	Q	T	E
chick CaBP I	Q	F	F	E	I	W	H	H	Y	D	S	D	G	N	G	Y	M	D	G	K	E	L	Q	N	F	I	Q	E	L
II	E	M	K	A	F	V	D	Q	Y	G	K	A	T	D	G	K	I	G	I	V	E	L	A	Q	V	L	P	T	E
III	D	F	M	Q	T	W	R	K	Y	D	S	D	H	S	G	F	I	D	S	E	E	L	K	S	F	L	K	D	L
IV	Y	T	E	I	M	L	R	M	F	D	A	N	N	D	G	K	L	E	L	T	E	L	A	R	L	L	P	V	Q
V	E	F	N	K	A	F	E	M	Y	D	Q	D	G	N	G	Y	I	D	E	N	E	L	D	A	L	L	K	D	L

Comparison of amino acid sequences of Ca binding regions of chick calmodulin, troponin C, chick CaBP and porcine CaBP. The top row represents the sequence identified from the structure of carp parvalbumin as binding Ca and specifies the constraints of each amino acid (25). E, glu; Z, a hydrophobic residue; O, oxygen-containing residue; G, gly; I, ile; - no amino-acid specified for this position. Sequences are represented using the single letter amino acid code. Ca binding domains within a molecule are distinguished by roman numerals. Chick CaBP domain I: positions 16-44, II, positions 58-86; III, positions 103-131; IV, positions 147-175; V, positions 191-219. Porcine and other mammalian CaBPs contain two amino acids A and N at positions 14 and 21 respectively which have to be looped out for their sequence to fit the consensus (27).

domains. More than half the nucleotides in domains I, III and V are identical and for the key amino acids involved in binding Ca about 70% of the nucleotides are homologous. In contrast domains II and IV are distinctly less homologous to the other three domains. In Table 4 the per

Table 3

Interdomain nucleotide homology in calcium-binding domains of chicken CaBP cDNA

Domains compared	Per Cent Homology
I:II	41 (52.0)
I:III	53 (71.0)
I:IV	37 (52.0)
I:V	52 (70.0)
II:IV	46 (48.0)

The assignment of the domains is based on the amino-acid sequence in Fig. 3 and is detailed in Table 2. The values in parenthesis are the per-cent homology for the constrained positions indicated in Table 2 by the symbols E, Z, O, G and I.

Table 4
Interdomain nucleotide homology for rat and chicken CaBP and chicken calmodulin cDNAs

Chicken CaBP	rat CaBP %		chicken calmodulin %	
domain	domain		domain	
	I	II	I	II
I	43 (51)	38 (57)	37 (50)	
II	46 (59)	37 (52)		42 (52)
III	37 (38)	41 (54)	38 (52)	
IV	37 (56)	42 (55)	37 (45)	42 (52)
V	43 (46)	44 (52)	50 (61)	

The assignment of the domains in chick CaBP is given in Table 2. For rat CaBP domain I is position 7-35 and domain II is 45-73, and for chick calmodulin domain I is 8-40 and domain III is 81-113 in their respective amino acid sequences. The values in parenthesis are the percent homology for the constrained positions indicated in Table 2 by the symbols E, Z, O, G and I.

cent homology between the five chicken domains and those in rat CaBP and chicken calmodulin are given. The homology is perhaps a little higher for the two chicken proteins than it is for the chicken and rat CaBP. In addition the chicken CaBP I, III and V domains are more homologous for rat CaBP II domain than I whereas domains II and IV tend to have a higher degree of homology with rat CaBP I domain, especially if only the most highly conserved amino acids are considered.

DISCUSSION

The use of a recently described method (12) for cDNA formation involving RNase H and DNA polymerase 1 for synthesis of the second strand produced ds transcripts of enriched intestinal mRNA of an average 2 Kb length. Previous estimates of the size of CaBP mRNA based on its position in a sucrose gradient and on the size of the polysomes synthesising CaBP indicated it to be about 2 Kb (8,10). After screening of recombinant colonies for inserts derived from mRNAs produced in response to 1,25-$(OH)_2$ D, two clones (pWH11 and pWH12) were identified containing all of the coding region for CaBP. pWH11 at 2.1 Kb was almost a full length copy of CaBP mRNA lacking little more than its poly A tract. This is the first example of the formation of such a clone for this physiologically important

protein. Previous reports described the formation of clones with inserts of <450 bp (5) and 150 bp (6) for chick intestinal CaBP and of 375 bp (4) for rat intestinal CaBP which were not full length copies of their respective mRNAs and did not contain all of the coding region.

An initiation codon ATG is located 101 nucleotides from the 5' end and establishes a coding sequence of 786 nucleotides sufficient to encode a protein of the size of CaBP. In common with many other cDNA sequences the 5' terminus has a high G + C content (78%) and has an abundance of CpG dinucleotides. It is probable that the first ATG triplet located at position 101 is the initiation codon. The sequence context of this ATG (CCATCATGA) is in good agreement with the proposed initiation codon consensus sequence CC(G/A)CCATGG (26). The sequence context of the second ATG (GGTACATGG) at position 194 differs significantly from the consensus and has an amino acid composition differing significantly from the measured values (23,24).

The nucleotide sequence of CaBP mRNA does not provide an explanation for its large size (2136 nucleotides) compared to the minimum required to encode CaBP (about 750 nucleotides). 1250 nucleotides are found at the non-translated 3' end of the sequence. Since the poly (A) tract was not present in pWH11 the poly-adenylation signal in the mRNA could not be identified. The untranslated 3' end is A+T rich and there are several examples of poly T of varying length up to a maximum of 16.

The homology observed between the Ca binding proteins extends to chick CaBP although the finding of five Ca binding sites in the putative CaBP sequence rather than the experimentally observed four sites was unexpected. Nevertheless this fifth site explains in part the relatively large size of CaBP compared to other proteins with four Ca binding sites such as calmodulin, tropinin C etc. The agreement with the consensus sequence for these sites is good for all the five sites but particularly for sites I, III and V. Indeed the agreement is better than for sites in brain S-100 protein and mammalian intestinal CaBP in which two additional amino acids are incorporated without deforming the site (27). This amino acid homology both between domains in the same protein and between proteins is reflected in the corresponding homology in the nucleotides from which they are derived. The nucleotide homology between the domains in chick CaBP and those in calmodulin indicates a high degree of conservation particularly for the constrained positions. For chick CaBP it seems that domains I, III and V are the most closely homologous and that II and IV are in a separate

homologous group. Consequently the Ca binding domains of chick CaBP as with other Ca binding proteins can be divided into two groups which may have arisen by gene duplication from a common ancestral gene.

Presumably post-translational modifications reduce the size of the initial translation product of CaBP mRNA to the observed value for the molecular weight and as a result the N-terminal amino acid is either part of or at the end of the Ca binding domain. In either case it is to be expected that the amino acids in this domain could not assume their proper configuration with the consequent interference in their Ca binding activity. Further work is necessary to establish the amino acid sequence of chick CaBP at least in the vicinity of the N-terminal residue. The mechanism by which the CaBP gene acquired its fifth calcium binding site may be revealed by knowledge of the structure of this gene obtained from genomic clones currently being isolated.

Acknowledgements

We thank Dr Celia Milstein for the supply of the vectors M13 mp18 and mp19. We also thank the MRC for financial support.

Abbreviation: CaBP, vitamin D-dependent chick calcium binding protein, M_r = 27500; 1,25-$(OH)_2$D, 1,25-dihydroxyvitamin D.

REFERENCES

1 Wasserman, R.H., Fullmer, C.S. & Taylor, A.N. (1978) In Vitamin D, Lawson, D.E.M. Ed., pp. 133-166. Acad. Press, London.
2 Hofmann, T., Kawakami, M., Hitchman, A.J.W., Harrison, J.E. & Dorington, K.J. (1979) Can. J. Biochem. 57, 737-748.
3 Lawson, D.E.M. (1978) in Vitamin D, Lawson, D.E.M. ed., pp 167-200. Acad Press, London.
4 Desplan, C., Thomasset, M. & Moukhtar, M. (1983) J. Biol. Chem. 258, 2762.
5 Hunziker, W., Siebert, P.D., King, M.W., Stucki, P., Dugaiczyk, A. & Norman, A.W. (1983) Proc. Natl. Acad. Sci. U.S.A. 80,4228-4232.
6 Ferrari, S., Battini, R., Leone, A., Ferrari, S., Torelli, G. & Barbiroli, B. (1984) Gene 30, 233-236.
7 Lawson, D.E.M. and Wilson, P.W. (1974) Biochem. J. 144, 573-583.
8 Emtage, J.S., Lawson, D.E.M. & Kodicek, E. (1974) Biochem. J. 140, 239-247.
9 Aviv, H. & Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
10 Spencer, R., Charman, M., Emtage, J.S. and Lawson, D.E.M. (1976) Eur. J. Biochem. 71, 399-409.
11 Pelham, H.R.B. and Jackson, R.J. (1976) Eur. J. Biochem. 67, 247-257.
12 Gubler, U. & Hoffman, B.J. (1983) Gene, 25, 263-269.
13 Norgard, M.V., Tocci, M.J. & Monahan, J.J. (1980) J. Biol. Chem. 255, 7665-7672.
14 Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.

15 Hanahan, D. & Meselson, M. (1983) Methods in Enz. 100, 333-342.
16 Maniatis, T., Frisch, E.F. & Sambrook, J. (1982) Molecular cloning: A Laboratory Manual. Cold Spring Harbour Laboratory, USA.
17 Parnes, J.R., Velan, B., Felsenfeld, A., Ramanathan, L., Ferrini, U., Appella, E. & Seidman, J.G. (1981) Proc. Natl. Acad. Sci. USA. 78, 2253-2257.
18 Paterson, B.M., Roberts, B.E. & Kuff, E.L. (1977) Proc. Natl. Acad. Sci. USA (74) 4370-4374.
19 Ricciardi, R.P., Miller, J.S. & Roberts, B.E. (1979) Proc. Natl. Acad. Sci. USA (76) 4927-4931.
20 Amersham M13 Cloning and sequencing handbook (1984) Amersham International.
21 Bankier, A.T. & Barrell, B.G. (1983) Nucleic Acid Biochemistry, B508, 1-34.
22 Biggin, M.D., Gibson, T.J. & Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA, 80, 3963-3965.
23 Fullmer, C.S. & Wasserman, R.H. (1975) Biochim. Biophys. Acta 393, 134-143.
24 Bredderman, P.B. & Wasserman, R.H. (1974) Biochemistry 13, 1687-1692.
25 Kretsinger, R.H. & Nockolds, C.E. (1973) J. Biol. Chem. 248, 3313-3326.
26 Kozak, M. (1984) Nucleic Acids Res. 12, 857-872.
27 Sjbenyi, D.M.E., Obendorf, S.K. & Moffat, K. (1981) Nature 294, 327-332.